SMS security on borrowed time

Two-factor authentication still has holes for fraudsters to slip through

BY STEPHEN WILSON

"Two-factor" authentication has come to mean many things in Internet banking. Originally it referred to a combination of something you know (a password or shared secret) and something you have (a credit card).

The pre-eminent example of two-factor authentication is of course the magnetic stripe card with PIN. The vulnerabilities of plastic cards to fraud, especially skimming, remind us of the important features of good two-factor solutions. Firstly, the physical factor must be difficult for an attacker to replicate, and secondly, users must still maintain a secret and obscure password, so that their token is protected against immediate abuse if lost.

For many years, online two-factor authentication was typified by one time password (OTP) generators: small, battery operated devices which display a pseudo-random logon code, either in response to a button push or automatically every half a minute or so. The user augments their regular account number and secret password with the OTP code to complete authentication.

One time passwords have been in long-time service for corporate remote logon, and have made their way slowly into Internet banking. It's difficult to tell definitively what has held up OTP in the market, but it's surely a combination of usability and cost (if not security). Institutions appear to have calculated the impact of OTP is high in return for the savings it should generate by reducing fraud and enhancing customer confidence.

In response to the cost of electronic tokens, a range of cheap "second factor" solutions have appeared, coinciding with what I believe to be a serious weakening in the definition of the term.

For instance, the Wikipedia entry on the topic doesn't mandate "something you have": "Two-factor authentication is a system wherein two different methods are used to authenticate. Using two factors as opposed to one delivers a higher level of authentication assurance". In a similar vein, the Australian Government Authentication Framework contains advice that "to guard against unjustified repudiation", businesses consider "periodically or randomly introducing a second e-authentication factor (for example, challenge-response using shared knowledge questions)".

Therefore security vendors have managed to promote as "two-factor" such variations as "matrix" or "grid" cards. These are printed look-up tables, unique to each customer, which furnish pseudo-random codes in response to row and column prompts. As such, they are really just a paper-based equivalent of electronic challenge-response calculators rolling out as an adjunct to EMV smartcards in Europe (see OBR June 2007). Crucially, matrix and grid cards can be duplicated by a thief and put back before the owner notices. It might therefore be more accurate to describe them as "one-and-a-bit factors", not two!

> **The vulnerabilities of plastic cards ... remind us of the important features of good two-factor solutions**

Yet over and above cost and inconvenience, the serious weakness in most two-factor authentication is vulnerability to the man-in-the-middle attack (see OBR June 2007 & April 2005). In the classic MITM attack, the transaction channel is taken over by an interloper who sets up a fake front door between the bank and the customer. The attacker lures the customer into initiating a transaction, but traps the account details and authentication codes before they can get to the server. They then compose their own fraudulent payment instruction, and forward that to the bank, together with a set of valid credentials from the hapless customer, including shared secrets and one time codes as applicable.

SMS authentication, as offered now by NAB, Commonwealth Bank and HSBC, provides improved resistance to MITM attack. Here the customer's mobile phone is utilised as the second physical factor. After initiating a transaction at the bank website, the server sends a one-time code back to the customer's phone, together with confirmation of the transfer amount and destination and the customer types the code back into the browser. If care is taken to match the details, they're protected against a MITM attack.

The logic of SMS authentication is that it's much harder to compromise the mobile phone network than the Internet. It's probably reasonable to assume that most customers have a mobile, and it's on their person most of the time. Another benefit is it may alert customers in the event that someone is tampering with their account.

But SMS was not designed to act as a second authentication factor, and it raises some serious issues. Firstly, there is no guarantee in the short message service standard that any SMS will ever arrive. If a banking confirmation code happens to never arrive, the inconvenience could be substantial. Moreover, help desks will have to find new ways to authenticate upset customers without creating "backdoor" security gaps.

Customers' mobile phone numbers also need to be kept secret, or else an attacker could send spoof messages (or even have a bot dial the phone) in order to fool or distract users amid transaction.

Above all, customers will need to read each SMS carefully to be sure the payment instruction is genuine. We know a substantial segment of the Internet banking market is vulnerable to phishing simply because they don't pay adequate attention to details[1]. It is this same segment that most needs two-factor authentication. SMS authentication is probably going to leave them vulnerable to frauds that exploit their credulity or naivety.

1 See e.g. http://cups.cs.cmu.edu/soups/2005/2005posters/15-wu.pdf and www.adambarth.com/papers/2007/jackson-simon-tan-barth.pdf.

Stephen Wilson is a leading international authority on identity management and information security. In early 2004, Stephen established Lockstep Consulting to provide independent security advice and to develop new smartcard solutions to identity theft.

swilson@lockstep.com.au